

# **ENSEMBLE-BASED ATMOSPHERIC DATA ASSIMILATION: A TUTORIAL**

Thomas M. Hamill

*University of Colorado and NOAA-CIRES Climate Diagnostics Center  
Boulder, Colorado, USA*

24 August 2004

*Corresponding author address:* Dr. Thomas M. Hamill, NOAA-CIRES Climate Diagnostics Center, R/CDC 1, 325 Broadway, Boulder, Colorado USA 80305-3328. tom.hamill@noaa.gov ; 1 (303) 497-3060 ; telefax 1 (303) 497-7013

## ABSTRACT

Ensemble-based data assimilation techniques are being explored as possible alternatives to current operational analysis techniques such as 3- or 4-dimensional variational assimilation. Ensemble-based assimilation techniques utilize an ensemble of parallel data assimilation and forecast cycles. The background-error covariances are estimated using the forecast ensemble and are used to produce an ensemble of analyses. The background-error covariances are flow dependent and often have very complicated structure, providing a very different adjustment to the observations than are seen from methods such as 3- dimensional variational assimilation. Though computationally expensive, ensemble-based techniques are relatively easy to code, since no adjoint nor tangent-linear models are required, and tests in simple models suggest that dramatic improvements over existing operational methods may be possible.

A review of the ensemble-based assimilation is provided here, starting from the basic concepts of Bayesian assimilation. Without some approximation, Bayesian assimilation is computationally impossible for large-dimensional systems. Assuming normality of error statistics and linearity of error growth, the state and its error covariance may be predicted optimally using Kalman filter (KF) techniques. The ensemble Kalman filter (EnKF) is then described. The EnKF is an approximation to the KF in that background-error covariances are estimated from a finite ensemble of forecasts. However, no assumptions about linearity of error growth are made. Recent algorithmic variants on the standard EnKF are also described, as well as methods for simplifying the computations and increasing the accuracy. Examples of ensemble-based assimilations are provided in simple and more realistic dynamical systems.

## 1. INTRODUCTION

The purpose of this article is to introduce the reader to promising new experimental methods for atmospheric data assimilation involving the use of ensemble forecasts (e.g., Evensen 1994, Evensen and van Leeuwen 1996, Houtekamer and Mitchell 1998, Burgers et al. 1998, Tippett et al. 2003, Anderson 2003, Evensen 2003, Lorenc 2003). There is a natural linkage between data assimilation and ensemble forecasting. Ensemble forecasts (Toth and Kalnay 1993, 1997, Molteni et al. 1996, Houtekamer et al. 1996a) are designed to estimate the flow-dependent uncertainty of the forecast; data assimilation techniques require accurate estimates of forecast uncertainty in order to optimally blend the prior forecast(s) with new observations. Ensemble-based assimilation methods integrate the two steps; the ensemble of forecasts is used to estimate forecast-error statistics during the data assimilation step, and the output of the assimilation is a set of analyses. This process is cycled, the short-term ensemble forecasts from the set of analyses providing the error statistics again for the next assimilation cycle.

Rather than starting with the specifics of recently proposed ensemble-based assimilation techniques, in this paper we will take a step back and try to motivate their use by quickly developing them from first principles, noting the approximations that have been made along the way. This will take us from Bayesian data assimilation (section 2), which is conceptually simple but computationally prohibitive, to the Kalman filter (section 3), a simplification assuming normality and linearity of error growth, to ensemble-based data assimilation methods (section 4), which may be more computationally tractable and robust. This discussion will include a description of stochastic and deterministic update algorithms, a simple example, and discussion of model error and covariance localization. Important ongoing research issues are discussed (section 5) and conclusions provided (section 6).

Several other useful review papers on ensemble-based data assimilation are available. Evensen (2003) provides a review of most of the proposed ensemble-based assimilation approaches, a more theoretical examination of the treatment of model errors, and a wide array of references to ensemble-based assimilation in the atmospheric and oceanographic literature. Lorenc (2003)

also reviews ensemble methods, and in particular provides some thoughts on the potential relative strengths and weaknesses compared to the current state-of-the-art assimilation algorithm, 4D-variational analysis (4D-Var). Tippett et al. (2003) discusses the similarities and differences between a number of the proposed algorithms, and Anderson (2003) discusses a way of interpreting ensemble-based techniques using simple linear regression terminology.

To keep the size of this manuscript manageable, several topics will be skipped. This manuscript will not describe the full variety of ensemble filters, in particular skipping a discussion of reduced-rank filters. Tangential subjects such as atmospheric predictability will be discussed only in relevance to the assimilation problem, and applications of ensemble filters to problems like adaptive observations will be neglected.

In subsequent discussion, the atmosphere state, which is of course a continuum, is assumed to be adequately described in discretized fashion, perhaps by the values of winds, temperature, humidity, and pressure at a set of grid points.

## 2. BAYESIAN DATA ASSIMILATION

Conceptually, the atmospheric data assimilation problem is a relatively simple one. The task at hand is to accurately estimate the probability density function (pdf) for the current atmospheric state given all current and past observations. Much of the material in this section follows Anderson and Anderson (1999). If the reader is interested in further background material on the subject, Lorenc (1986) provides a formulation of data assimilation in a Bayesian context, and Cohn (1997) provides a more rigorous statistical formulation of the problem.

When considering Bayesian assimilation, there are two general steps to the assimilation. Assume that a pdf of the state of the atmosphere is available (in the lack of any knowledge, this may be the climatological pdf). The first step is to assimilate recent observations, thereby sharpening the pdf. The second step is to propagate the pdf forward in time until new observations are available. If the pdf is initially sharp (i.e., the distribution is relatively narrow), then chaotic dynamics and model uncertainty will usually broaden the probability distribution. The update

and forecast steps are then repeated. We will describe each of these steps separately, starting with the assimilation of new observations.

*a. Bayesian updating*

Assume that an estimate of the pdf has been propagated forward to a time when observations are available. The state can be estimated more specifically by incorporating information from the new observations. This will be termed the “update.”

The following notational convention is used. Boldface characters will denote vectors or matrices, while use of the italicized font denotes a scalar.  $\mathbf{x}_{t-1}^t$  will denote the  $n$ -dimensional true model state at time  $t - 1$ :  $\mathbf{x}_{t-1}^t = [x_{t-1(1)}^t, \dots, x_{t-1(n)}^t]$ . Also, assume a collection of observations  $\psi_t$ . This vector includes observations  $\mathbf{y}_t$  at the most recent time as well as observations at all previous times  $\psi_t = [\mathbf{y}_t, \psi_{t-1}]$ , where  $\psi_{t-1} = [\mathbf{y}_{t-1}, \dots, \mathbf{y}_0]$ . There are  $M_t$  observations at time  $t$ , i.e.,  $\mathbf{y}_t = [y_{t(1)}, \dots, y_{t(M_t)}]$ . Let  $P(\mathbf{x}_t^t)$  be a multivariate probability density function, defined such that  $Pr(\mathbf{a} \leq \mathbf{x}_t^t \leq \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} P(\mathbf{x}_t^t) d\mathbf{x}_t^t$ , and probability density integrates to 1.0 over the entire phase space.

Formally, the update problem is to accurately estimate  $P(\mathbf{x}_t^t | \psi_t)$ , the probability density estimate of the current atmospheric state given the current and past observations. Bayes’ Rule tells us that this quantity can be re-expressed as

$$P(\mathbf{x}_t^t | \psi_t) \propto P(\psi_t | \mathbf{x}_t^t) P(\mathbf{x}_t^t). \quad (1)$$

Bayes’ Rule is usually expressed with a normalization constant in the denominator on the right-hand side of (1); for simplicity, the term in the denominator will be dropped here, and it is assumed that when coded, the developer will ensure that probability density integrates to 1.0.

One hopefully minor assumption is made: observation errors are independent from one time to the next. Hence,  $P(\psi_t | \mathbf{x}_t^t) = P(\mathbf{y}_t | \mathbf{x}_t^t) P(\psi_{t-1} | \mathbf{x}_t^t)$ . This may not be true for observations from satellites, where instrumentation biases may be difficult to remove. Also, errors of observation representativeness (Daley 1993) may be flow dependent and correlated in time. But under

this assumption, (1) is equivalent to

$$P(\mathbf{x}_t^t | \psi_t) \propto P(\mathbf{y}_t | \mathbf{x}_t^t) P(\psi_{t-1} | \mathbf{x}_t^t) P(\mathbf{x}_t^t). \quad (2)$$

By Bayes' Rule again,  $P(\psi_{t-1} | \mathbf{x}_t^t) P(\mathbf{x}_t^t) \propto P(\mathbf{x}_t^t | \psi_{t-1})$ . Hence, (2) simplifies to

$$P(\mathbf{x}_t^t | \psi_t) \propto P(\mathbf{y}_t | \mathbf{x}_t^t) P(\mathbf{x}_t^t | \psi_{t-1}). \quad (3)$$

In principle, equation (3) is elegantly simple. It expresses a recursive relationship: the “posterior,” the pdf for the current model state given all the observations, is a product of the the probability distribution for the current observations  $P(\mathbf{y}_t | \mathbf{x}_t^t)$  and the “prior,”  $P(\mathbf{x}_t^t | \psi_{t-1})$ , also known as the “background.” The prior is the pdf of the model state at time  $t$  given all the past observations up to time  $t - 1$ . Typically, the prior will have been estimated in some fashion from a cycle of previous data assimilations and short-term forecasts up to the current time; approximations of how this may be computed will be discussed in section 2b.

Let's now demonstrate the update step of Bayesian assimilation with a simple example.  $P(\mathbf{x}_t^t | \psi_{t-1})$  is an estimate of the prior for a two-dimensional model state. This was produced by assimilating all prior observations up to and including time  $t - 1$  and estimating in some manner how that pdf has evolved in the time interval between  $t - 1$  and  $t$ . Consider how to update the pdf given a new scalar observation  $y$ , which in this example is observing the same quantity as the first component of the state vector measures. The pdf for the observation  $P(y_t | \mathbf{x}_t^t)$  is assumed to be distributed normally about the actual observation,  $\sim N(y_t, \sigma^2)$ . Here, let  $y_t = 58$  and  $\sigma^2 = 100$ .

Selected contours of the prior are plotted in Fig. 1a ; as shown, the prior is bimodal. The shape of the marginal prior distributions  $P(x_{t(1)} | \psi_{t-1})$  and  $P(x_{t(2)} | \psi_{t-1})$  are plotted along each axis in solid lines. The dashed line denotes the observation probability distribution  $P(y_t | \mathbf{x}_t^t)$ . This probability varies with the value  $x_{t(1)}$ , but given  $x_{t(1)}$  is the same for any value of  $x_{t(2)}$ . The updated posterior distribution is computed using (3) and is shown in Fig. 1b. Note that the assimilation of the observation enhanced the probability in the lobe overlapping the observation

distribution and decreased it in the other lobe. Overall, the posterior distribution is more sharp (specific) than the prior, as is expected.

#### *b. Forecasting of probability density*

Assume that we have an (imperfect) nonlinear forecast model operator  $\mathcal{M}$  so that the time evolution of the state can be written as a stochastic differential equation:

$$d\mathbf{x}_t^t = \mathcal{M}(\mathbf{x}_t^t) dt + \mathcal{G}(\mathbf{x}_t^t) d\mathbf{q}. \quad (4)$$

$d\mathbf{q}$  is a Brownian-motion process with covariance  $\mathbf{Q}_t dt$  and  $\mathcal{G}$  is the model-error forcing. Conceptually, the time evolution of the pdf can be modeled with the Fokker-Planck equation (e.g., Gardiner 1985, section 5.3):

$$\frac{\partial P(\mathbf{x}_t^t)}{\partial t} = -\nabla \cdot \left[ \mathcal{M}(\mathbf{x}_t^t) P(\mathbf{x}_t^t) \right] + \sum_{i,j} \frac{\partial^2}{\partial x_{t(i)}^t \partial x_{t(j)}^t} \left( \frac{\mathcal{G} \mathbf{Q}_t \mathcal{G}^T}{2} \right)_{ij} P(\mathbf{x}_t^t) \quad (5)$$

If  $\mathcal{G} \mathbf{Q}_t \mathcal{G}^T$  is zero, then only the first term remains, and the Fokker-Planck equation reduces to the Liouville equation (Ehrendorfer 1994ab), a continuity equation for the conservation of probability. Probability thus diffuses with time according to the chaotic dynamics of the forecast model. The second term includes the effects of model error, including the increased diffusion of probability due to model uncertainty as well as noise-induced drift (Sardeshmukh et al. 2001).

#### *c. Limitations of Bayesian data assimilation*

Unfortunately, neither the update nor the forecast steps in Bayesian data assimilation can be applied directly to real-world NWP applications. For the update step, one problem with modeling a complicated pdf in higher dimensions is the “curse of dimensionality” (e.g., Hastie et al. 2001, pp. 22-27). Were one to try estimate the probability density in a higher-dimensional space using a small ensemble, one would find that the model of probability was very poor unless simplifying assumptions about the form of the distribution were made. Even were this problem surmountable, the computational cost would be extravagant. In the prior example the probability density was evaluated on a  $100 \times 100$  grid. Suppose a similarly complicated structure for the

prior existed in 100 dimensions. Then if the joint probabilities were monitored on a similar grid for each dimension, this would involve evaluating and modifying  $100^{100}$  density estimates. Such computations are already prohibitive for a 100-dimensional model state; the problem becomes incomprehensible for model states of  $O(10^7)$ .

Similarly, the Fokker-Planck equation is conceptually appealing but practically unusable for large systems; The pdf cannot be forecast in high-dimensional systems using (5) due to computational constraints. Consequently, Monte-Carlo techniques are typically applied. Assume that it is possible to randomly sample  $P(\mathbf{x}_t^l \mid \psi_t)$ . Then an ensemble of forecasts is conducted forward from the random sample, including some appropriate way to include the uncertainty due to model error. In this manner, a random sample of state estimates can be produced up to the next time observational data is available. In section 4, we will return to consider how to utilize the ensemble forecast data in the update process.

### 3. KALMAN FILTERS

An approximation to Bayesian state estimation is now considered under assumptions of linearity of error growth and normality of error distributions. This approximate method is known as the *Kalman filter* (Kalman 1960, Kalman and Bucy 1961, Jazwinski 1970 sect. 7.3, Gelb 1974 sect. 4.2, Maybeck 1979 sect 5.3, Ghil 1989, Daley 1991 sect 13.3, Cohn 1997, Talagrand 1997, Daley 1997). There are two components of the Kalman filter, an update step where the state estimate and an estimate of the forecast uncertainty are adjusted to new observations, and a forecast step, where the updated state and the uncertainty estimate are propagated forward to the time when the next set of observations become available.

#### *a. The extended Kalman filter*

We now consider an implementation of the Kalman filter called the extended Kalman filter, or “EKF” (Jazwinski 1970, Gelb 1974, Ghil and Malanotte-Rizzoli 1991, Gauthier et al. 1993, Bouttier 1994). The EKF assumes that background and observation error distributions are Gaussian:  $\mathbf{x}_t^b = \mathbf{x}_t^l + e$ , where  $e \sim N(0, \mathbf{P}_t^b)$ . That is, the probability density of the prior is dis-



tributed as a multivariate normal distribution with known  $n \times 1$  mean background  $\mathbf{x}_t^b$  and  $n \times n$  background-error covariance matrix  $\mathbf{P}_t^b$ . Similarly,  $\mathbf{y} = \mathcal{H}(\mathbf{x}_t^b) + \epsilon$ , where  $\epsilon \sim N(0, \mathbf{R})$  and  $\mathcal{H}$  is the  $M_t \times n$  “forward” operator that maps the state to the observations.  $\mathbf{H}$  is the  $m \times n$  Jacobian matrix of  $\mathcal{H} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}}$ . Also, let  $\mathcal{M}$  represent the nonlinear model forecast operator.  $\mathbf{M}$  is the  $n \times n$  Jacobian matrix of  $\mathcal{M}$ ,  $\mathbf{M} = \frac{\partial \mathcal{M}}{\partial \mathbf{x}}$ .  $\mathbf{M}$  is often called the *transition matrix* between times  $t$  and  $t + 1$ .  $\mathbf{M}^T$  is its adjoint (see Le Dimet and Talagrand (1986) and Lacarra and Talagrand (1988)).  $\mathbf{Q}$  will represent the  $n \times n$  covariance of model errors accumulated between update cycles.

The EKF equations are

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}(\mathbf{y}_t - \mathcal{H}(\mathbf{x}_t^b)), \quad (6a)$$

$$\mathbf{K} = \mathbf{P}_t^b \mathbf{H}^T (\mathbf{H} \mathbf{P}_t^b \mathbf{H}^T + \mathbf{R})^{-1}. \quad (6b)$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}_t^b \quad (6c)$$

$$\mathbf{x}_{t+1}^b = \mathcal{M}(\mathbf{x}_t^a) \quad (6d)$$

$$\mathbf{P}_{t+1}^b = \mathbf{M} \mathbf{P}_t^a \mathbf{M}^T + \mathbf{Q} = \mathbf{M} (\mathbf{M} \mathbf{P}_t^a)^T + \mathbf{Q}. \quad (6e)$$

Equations (6a-c) describe the update step. The optimal analysis state  $\mathbf{x}_t^a$  is estimated by correcting the background  $\mathbf{x}_t^b$  toward the “observation increment”  $\mathbf{y}_t - \mathcal{H}(\mathbf{x}_t^b)$ , weighted by the Kalman-gain matrix  $\mathbf{K}$ . The effect of  $\mathbf{K}$  is to apply observation increments to correct the background at relevant surrounding grid points. Equation (6c) indicates how to update the background-error covariance to reflect the reduction in uncertainty from assimilating the observations. Equations (6d-e) propagate the resulting analysis and error covariance forward in time to when observations are next available. The expected analysis state is propagated forward with the full nonlinear forecast model. Model errors are assumed to be uncorrelated with the growth of analysis errors through the tangent-linear forecast dynamics.

The appeal of the Kalman filter relative to an analysis scheme like three-dimensional variational assimilation (3D-Var; Lorenc 1986, Parrish and Derber 1992) is that the error covariances of the analysis and the background evolve with time. The analysis reduces error variances in locations where accurate observations are plentiful, and the error covariances are also forecast for-

ward in time, growing at a rate proportional to the local error growth. Consequently, the structure of the background-error covariances and hence the adjustment to observations  $\mathbf{x}_t^a - \mathbf{x}_t^b$  can be quite complicated and flow- and time-dependent (e.g., Bouttier 1994).

*d. Considerations in the use of Kalman filters*

What approximations may limit the accuracy of the EKF? First, Kalman filters assume linear growth and normality of errors, for the assimilation problem becomes much more tractable when these assumptions are made. Non-normality of the prior such as the bimodality in Fig. 1a is typically assumed to be uncommon in atmospheric data assimilation. These linear and normal assumptions may be inappropriate for atmospheric data assimilations of moisture, cloud cover, and other aspects of the model state that may be very sensitive to motions at small scales, where the time scale of predictability is small and errors grow and saturate rapidly. Similarly, if observations are not regularly available, error covariances estimated with tangent-linear dynamics may grow rapidly without bound (Evensen 1992, Gauthier et al. 1993, Bouttier 1994).

Second, error statistics must be carefully estimated and monitored; in particular, it is important that the background-error covariance matrix be estimated accurately. For example, if background error variances are underestimated, the EKF will assume the error statistics are indicating that the background is relatively more accurate than the nearby observations and thus will not correct the background to the observations to the extent it should (Daley 1991, p. 382). Estimating  $\mathbf{Q}$  may be particularly difficult. In practice, accurately determining even the time-averaged statistics of  $\mathbf{Q}$  may be quite complicated (Cohn and Parrish 1991, Daley 1992, Dee 1995, Blanchet et al. 1997). For both the Kalman filter and ensemble-based methods, the accuracy of the assimilation is likely to strongly depend on this assumed model for  $\mathbf{Q}$ . Methods for estimating  $\mathbf{Q}$  will be discussed for ensemble-based methods in section 4d.

Another disadvantage of the Kalman filters for atmospheric data assimilation is the computational expense. Though Kalman filters provide a dramatic reduction in the computational cost relative to full Bayesian data assimilation, for a highly dimensional state vector, the computational costs in weather prediction models may still be impossibly large. Consider the last line in

(10). For an  $n$ -dimensional model state vector, it will require  $2n$  applications of  $\mathbf{M}$  to forecast the error covariances. Some reductions of computational expense may be possible. For example, there have been suggestions that this computation may be more practical if the tangent-linear calculations are performed in a subspace of the leading singular vectors (Fisher 1998, Farrell and Ioannou 2001).

Much more can be said about the Kalman filter, such as its equivalence to 4D-Var under certain assumptions (Li and Navon 2001), the manner of computing  $\mathbf{M}$ , iterated extensions of the basic extended Kalman filter (Jazwinski 1970, Gelb 1974, Cohn 1997), and the properties of its estimators (which, in the case of the discrete filter, if assumptions hold, provide the Best Linear Unbiased Estimate, or BLUE; see Talagrand 1997).

#### 4. ENSEMBLE-BASED DATA ASSIMILATION

Ensemble-based assimilation algorithms may be able to provide a more accurate analysis in situations where nonlinearity is strong and statistics exhibit some non-normality. If these assimilation algorithms can work accurately with many fewer ensemble members than elements in the state vector, then they will be computationally much less expensive as well. Consequently, many researchers have proposed a variety of ensemble-based assimilation methods. Despite the many differences between the various ensemble-based algorithms, all are comprised of a finite number (perhaps 10 to a few hundred) parallel data assimilation and short-range forecast cycles. Background-error covariances are modeled using the ensemble of forecasts, and an ensemble of analyses are produced, followed by an ensemble of short-term forecasts to the next time observations are available. Ensemble-based assimilation algorithms also have the desirable property that if error dynamics are indeed linear and the error statistics Gaussian, then as the ensemble size increases, the state and covariance estimate from ensemble algorithms converge to those obtained from the extended Kalman filter (Burgers et al. 1998).

The concepts behind ensemble assimilation methods have been used in engineering and aerospace applications as far back as the 1960's (Potter 1964, Andrews 1968, Kaminski et al.

1971, Maybeck 1979, ch. 7). Leith (1983) sketched the basic idea for atmospheric data assimilation. The idea was more completely described and tested in an oceanographic application by Evensen (1994) and in atmospheric data assimilation by Houtekamer and Mitchell (1998). Here, two classes of ensemble-based filters will be described, stochastic and deterministic. Both classes propagate the ensemble of analyses with nonlinear forecast models; the primary difference is whether or not random noise is applied during the update step.

For notational simplicity, the  $t$  time subscript used in previous sections is dropped; it is assumed unless noted otherwise that we are interested in estimating the state pdf at time  $t$ . We start off by assuming that we have an ensemble of forecasts that randomly sample the model background errors at time  $t$ . Let's denote this ensemble as  $\mathbf{X}^b$ , a matrix whose columns are comprised of ensemble member's state vectors:

$$\mathbf{X}^b = (\mathbf{x}_1^b, \dots, \mathbf{x}_m^b), \quad (7)$$

The subscript now denotes the ensemble member. The ensemble mean  $\bar{\mathbf{x}}^b$  is defined as

$$\bar{\mathbf{x}}^b = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^b. \quad (8)$$

The perturbation from the mean for the  $i$ th member is  $\mathbf{x}_i'^b = \mathbf{x}_i^b - \bar{\mathbf{x}}^b$ . Define  $\mathbf{X}'^b$  as a matrix formed from an ensemble of perturbations

$$\mathbf{X}'^b = (\mathbf{x}_1'^b, \dots, \mathbf{x}_m'^b) \quad (9)$$

and let  $\hat{\mathbf{P}}^b$  represent an estimate of  $\mathbf{P}^b$  from a finite ensemble

$$\hat{\mathbf{P}}^b = \frac{1}{m-1} \mathbf{X}'^b \mathbf{X}'^{bT}. \quad (10)$$

Unlike the Kalman filter or 3D-Var, the background-error covariance estimate is generated from a specially constructed ensemble of nonlinear forecasts. The finite sample will introduce errors (see, e.g., Casella and Berger, sect. 5.4 and Hamill et al. 2001, sect. 2) relative to the EKF. However, estimating the covariances using an ensemble of nonlinear model forecasts may provide a powerful advantage over the EKF. Envision a situation where errors grow rapidly but saturate at low amplitude; the linear assumption of error growth in the EKF will result in an overestimate of background error variance, but the differences among ensemble members will not grow

without bound and thus should provide a more accurate model of the actual background-error statistics. Further, unlike most 3D-Var implementations, the background-error covariances can vary in time and space. If this error covariance model is relatively accurate, it will thus provide a better adjustment to the observations.

Figure 2 illustrates the potential benefit from estimating background-error covariances using an ensemble-based data assimilation system. Here we see a snapshot of sea-level pressure background-error covariances with five locations around the Northern Hemisphere, estimated from a 100-member ensemble. The data was taken from the ensemble data assimilation experiment of Whitaker et al. (2004), which tested the efficacy of assimilating only a sparse network of surface pressure observations concentrated over the U.S., Europe, and east Asia. A covariance localization with a correlation length of approximately  $\sim 2700$  km was applied (see section 4e). Notice that the magnitude and the spatial structure of the background-error covariances changes from one location to the next, with larger covariances for the point south of Alaska and northern Russia and smaller covariances at other locations. The horizontal extent of the positive covariance also changed markedly from one location to the next. The background-error covariances control the magnitude of the adjustment to the observation, drawing more to observations when background errors are large. Hence, observations will affect the analysis very differently around each of the five locations, which is the essence of why ensemble-based algorithms may outperform other methods.

We will first consider the update step in two general classes of ensemble filters, stochastic and deterministic. A discussion of how forecast errors are estimated and model error treated then follows. A brief example of the update step is then provided, followed by a description of an important algorithmic modification, covariance localization.

#### *a. Stochastic update algorithms*

The most well-known stochastic ensemble-based data assimilation algorithm is the *ensemble Kalman filter*, or “EnKF” (Houtekamer and Mitchell 1998, 1999, 2001, Burgers et al. 1998, Keppenne 2000, Mitchell and Houtekamer 2000, Hamill et al. 2001, Keppenne and Rienecker

2002, Mitchell et al. 2002, Hamill and Snyder 2002, Houtekamer et al. 2004). This algorithm updates each member to a different set of observations perturbed with random noise. Because randomness is introduced every assimilation cycle, the update is considered stochastic.

The EnKF performs an ensemble of parallel data assimilation cycles,  $i = 1, \dots, m$ , with each member updated to a somewhat different realization of the observations:

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \hat{\mathbf{K}}(\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i^b)). \quad (11)$$

In (11), the  $\mathbf{y}_i = \mathbf{y} + \mathbf{y}'_i$  are “perturbed observations,” defined such that  $\mathbf{y}'_i \sim N(0, \mathbf{R})$ , and

$$\frac{1}{m} \sum_{i=1}^m \mathbf{y}'_i = 0. \quad (12)$$

The  $m$  sets of perturbed observations are thus created to update the  $m$  different background fields. Here, in (11),

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}^b \mathcal{H}^T (\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T + \mathbf{R})^{-1}, \quad (13)$$

similar to the Kalman gain of the EKF gain in (6b), but permitting a possibly nonlinear  $\mathcal{H}$  and using the ensemble to estimate the background-error covariance matrix as in (10).

Notice that the EnKF assimilates perturbed observations in (11) rather than the observations themselves. To understand this, let  $\mathbf{X}^{/a}$  be a matrix of analysis ensemble member deviations from the analysis mean state, as (9) defined background deviations. Let  $\hat{\mathbf{P}}^a$  be formed from the ensemble of analyses assimilating perturbed observations using (11). Then as the ensemble size approaches infinity and if the dynamics are linear,  $\hat{\mathbf{P}}^a = \frac{1}{m-1} \mathbf{X}^{/a} \mathbf{X}^{/aT} \rightarrow \mathbf{P}^a$ , where  $\mathbf{P}^a$  is the extended Kalman filter analysis-error covariance from (6c) (Burgers et al. 1998). If unperturbed observations are assimilated in (11) without other modifications to the algorithm, the analysis-error covariance will be underestimated, and observations will not be properly weighted in subsequent assimilation cycles.

Adding noise to the observations in the EnKF can introduce spurious observation-background error correlations that can bias the analysis-error covariances, especially when the ensemble size is small (Whitaker and Hamill 2002). Pham (2001) proposed an alternative to perturbing the ob-

servations, adding noise to background forecasts in a manner that also ensures analysis-error covariances are equal to those produced by the EKF. Anderson (2003) proposed a sequential observation processing method that minimizes this effect. Houtekamer and Mitchell (1998) proposed the use of a “double” EnKF with two parallel sets of ensembles, each set used to estimate background-error covariances to update the other set. See van Leeuwen (1999), Houtekamer and Mitchell (1999), and Whitaker and Hamill (2002) for a discussion of covariance biases in the single and double EnKFs.

Several algorithms have been proposed for simplifying and parallelizing the coding of the EnKF. One technique that is uniformly used is to form the Kalman gain from the ensemble without ever forming the actual background-error covariance matrix. For a complex numerical weather prediction model with a high-dimensional state vector, explicitly forming  $\hat{\mathbf{P}}^b$  as in (10) would be computationally prohibitive; for example, in a model with  $10^7$  elements in its state, storing and readily accessing the  $10^{14}$  elements of  $\hat{\mathbf{P}}^b$  is not possible. However, in ensemble-based methods,  $\hat{\mathbf{K}}$  can be formed without ever explicitly computing the full  $\hat{\mathbf{P}}^b$  (Evensen 1994, Houtekamer and Mitchell 1998). Instead, the components of  $\hat{\mathbf{P}}^b \mathcal{H}^T$  and  $\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T$  of  $\hat{\mathbf{K}}$  are computed separately. Define

$$\overline{\mathcal{H}(\mathbf{x}^b)} = \frac{1}{m} \sum_{i=1}^m \mathcal{H}(\mathbf{x}_i^b),$$

which represents the mean of the estimate of the observation interpolated from the background forecasts. Then

$$\hat{\mathbf{P}}^b \mathcal{H}^T = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{x}_i^b - \overline{\mathbf{x}^b}) (\mathcal{H}(\mathbf{x}_i^b) - \overline{\mathcal{H}(\mathbf{x}^b)})^T, \quad (14)$$

and

$$\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T = \frac{1}{m-1} \sum_{i=1}^m (\mathcal{H}(\mathbf{x}_i^b) - \overline{\mathcal{H}(\mathbf{x}^b)}) (\mathcal{H}(\mathbf{x}_i^b) - \overline{\mathcal{H}(\mathbf{x}^b)})^T. \quad (15)$$

Of course, if the number of observations is as large as the elements in the model state,  $\hat{\mathbf{P}}^b \mathcal{H}^T$  and  $\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T$  will be as large as  $\hat{\mathbf{P}}^b$ , negating this advantage. However, another possible coding simplification is *serial processing*. If observations have independent errors uncorrelated with the background, they can be assimilated simultaneously or serially (sequentially), producing

the same result (Kaminski et al. 1971, Gelb 1974 p. 304, Bishop et al. 2001). The analysis ensemble after the assimilation of the first observation is used as the background ensemble for the assimilation of the second, and so on. When observations are assimilated serially, for each observation that is assimilated,  $\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T$  and  $\mathbf{R}$  become scalars. Thus, the inverse  $(\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T + \mathbf{R})^{-1}$  in the gain matrix is trivial to compute. Also, the application of the covariance localization, discussed later, is much more straightforward to apply. Serial stochastic ensemble filters have been demonstrated in Houtekamer and Mitchell (2001), Hamill et al. (2001), Hamill and Snyder (2002), and Anderson (2003).

The equivalence of serial and simultaneous processing is only true if observations have independent errors (Kaminski et al. 1971). Practically, however, many observations may have vertically or horizontally correlated errors. Consider two alternatives to deal with this. First, if the size of a batch of observations with correlated errors is relatively small, these batches can be processed simultaneously without much more computational expense (Houtekamer and Mitchell 2001, Mitchell et al. 2002, Houtekamer et al. 2004); the matrix inverse of  $(\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T + \mathbf{R})^{-1}$  should not be prohibitively expensive. Another option is to transform the observations and the forward operator so that the observations are effectively independent. The method for doing this is derived in the Appendix.

Several investigators have proposed speeding up the performance of the stochastic EnKF by separately updating different grid points independently on different processors. Keppenne and Rienecker (2002) designed an algorithm whereby all observations in the region of a particular set of grid points are simultaneously assimilated to update those grid points. Other distinct sets of grid points are updated independently. Houtekamer and Mitchell (2001) propose a method that uses both serial processing of observations and processing different regions separately from one another. They also discuss other ways of minimizing the amount of information that needs to be swapped between processors on a parallel computer. Reichle et al. (2002ab) and Reichle and Koster (2003) demonstrate a parallelized EnKF algorithm applied to soil-moisture state estimation.



### *b. Deterministic update algorithms*

Several methods have been proposed to correct the background ensemble to new observations so that  $\hat{\mathbf{P}}^a \rightarrow \mathbf{P}^a$  without adding random noise. Algorithms that do not add stochastic noise are called *deterministic* algorithms, so named because if the background ensemble and the associated error statistics are known, the ensemble of analysis states will be completely known as well. These algorithms (e.g., Lermusiaux and Robinson 1999, Bishop et al. 2001, Anderson 2001, Whitaker and Hamill 2002, Lermusiaux 2002, Ott et al. 2003) update in a way that generates the same analysis-error covariance update that would be obtained from the Kalman filter, assuming that the Kalman filter's background-error covariance is modeled from the background ensemble. Tippett et al. (2003) describes the similarities and differences between several of these algorithms. In each, the background-error covariances are never explicitly formed, with manipulations being performed using the matrix square root (i.e., eq (9), the matrix of ensemble member deviations from the mean). As pointed out in Tippett et al., since  $\hat{\mathbf{P}}^b = \frac{1}{m-1} \mathbf{X}^b \mathbf{X}^{bT}$ , given a matrix  $\mathbf{U}$  representing any  $n \times n$  orthogonal transformation such that  $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$ , then  $\hat{\mathbf{P}}^b$  can also be represented as  $\hat{\mathbf{P}}^b = \frac{1}{m-1} (\mathbf{X}^b \mathbf{U}) (\mathbf{X}^b \mathbf{U})^T$ . Hence, many square-root filters can be formulated that produce the same analysis-error covariance.

Since Tippett (2003) reviews many of these methods, we will explicitly describe only one of these, a particularly simple implementation, the “ensemble square-root filter,” or “EnSRF,” described by Whitaker and Hamill (2002), which is mathematically equivalent to the filter described in Anderson (2001). The EnSRF algorithm has been used for the assimilation at the scale of thunderstorms by Snyder and Zhang (2003), Zhang et al. (2003) and Dowell et al. (2004). Whitaker et al. (2004) used the algorithm in for the global data assimilation of surface pressure observations. Like the EnKF, the EnSRF conducts a set of parallel data assimilation cycles. It is convenient in the EnSRF to update the equations for the ensemble mean (denoted by an overbar) and the deviation of the  $i$ th member from the mean separately:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \hat{\mathbf{K}} \left( \mathbf{y} - \mathcal{H}(\bar{\mathbf{x}}^b) \right), \quad (16)$$

$$\mathbf{x}_i'^a = \mathbf{x}_i'^b - \tilde{\mathbf{K}}\mathcal{H}(\mathbf{x}_i'^b). \quad (17)$$

Here,  $\hat{\mathbf{K}}$  is the traditional Kalman gain as in Eq. (13), and  $\tilde{\mathbf{K}}$  is the “reduced” gain used to update deviations from the ensemble mean.

When sequentially processing independent observations,  $\hat{\mathbf{K}}$ ,  $\tilde{\mathbf{K}}$ ,  $\mathcal{H}\hat{\mathbf{P}}^b$  and  $\hat{\mathbf{P}}^b\mathcal{H}^T$  are all  $n$ -dimensional vectors, and  $\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T$  and  $\mathbf{R}$  are scalars. Thus, as first noted by Potter (1964), when observations are processed one at a time,

$$\tilde{\mathbf{K}} = \left( 1 + \sqrt{\frac{\mathbf{R}}{\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T + \mathbf{R}}} \right)^{-1} \hat{\mathbf{K}}. \quad (18)$$

The quantity multiplying  $\hat{\mathbf{K}}$  in Eq. (18) thus becomes a scalar between 0 and 1. This means that, in order to obtain the correct analysis-error covariance with unperturbed observations, a modified Kalman gain that is reduced in magnitude relative to the traditional Kalman gain is used to update deviations from the ensemble mean. Consequently, deviations from the mean are reduced less in the analysis using  $\tilde{\mathbf{K}}$  than they would be using  $\hat{\mathbf{K}}$ . In the stochastic EnKF, the excess variance reduction caused by using  $\hat{\mathbf{K}}$  to update deviations from the mean is compensated for by the introduction of noise to the observations.

In the EnSRF, the mean and departures from the mean are updated independently according to Eqs. (16) and (17). If observations are processed one at a time, the EnSRF requires about the same computation as the traditional EnKF with perturbed observations, but for moderately sized ensembles and processes that are generally linear and Gaussian, the EnSRF produces analyses with significantly less error (Whitaker and Hamill 2002). Conversely, Lawson and Hansen (2003) suggest that if multimodality is typical and ensemble size is large, the EnKF will perform better.

Another deterministic update algorithm is the Ensemble Transform Kalman filter (ETKF) of Bishop et al. (2001). The ETKF finds the transformation matrix  $\mathbf{T}$  such that  $\hat{\mathbf{P}}^a = \frac{1}{m-1} (\mathbf{X}^b\mathbf{T}) (\mathbf{X}^b\mathbf{T})^T \rightarrow \mathbf{P}^a$ . (see Bishop et al. for details on the computation of  $\mathbf{T}$ ). Compared with the EnSRF, an advantage of the ETKF is its computational speed; a disadvantage is that the ETKF cannot apply covariance localizations (section 5), which may make the analyses very inaccurate unless large

ensembles are used. The ETKF has been successfully demonstrated for generating perturbed initial conditions for ensemble forecasts about a mean state updated using 3D-Var (Wang and Bishop 2003), and hybrid ETKF-variational schemes are being explored (Etherton and Bishop 2003).

Another deterministic update algorithm was proposed by Ott et al. (2003). This algorithm proposes processing individual grid points independently in a manner similar to that proposed in Keppenne and Rienecker (2002) and Evensen (2003, section 4.4). However, like the algorithms discussed above, it adjusts the ensemble to provide the analytically correct analysis-error covariance without perturbing the observations.

*c. A simple demonstration of stochastic and deterministic update steps*

Consider again the Bayesian data assimilation problem illustrated in Fig. 1. There, a bimodal 2-D probability distribution was updated to an observation of one component. Let's explore the characteristics of the EnKF and EnSRF update applied to this problem.

A 100-member random sample was first generated from the bimodal pdf in Fig. 1a. These samples are denoted by the black dots in Fig. 3a. Let's keep track of the assimilation for one particular member, denoted by the larger black dot.

The EnKF and EnSRF adjust the background to the observations with weighting factors that assume the distributions are normal. Estimated from this random sample, the background-error covariance is

$$\hat{\mathbf{P}}^b = \begin{pmatrix} \sigma^2(x_{(1)}^b) & Cov(x_{(1)}^b, x_{(2)}^b) \\ Cov(x_{(1)}^b, x_{(2)}^b) & \sigma^2(x_{(2)}^b) \end{pmatrix} \simeq \begin{pmatrix} 150.73 & 109.70 \\ 109.70 & 203.64 \end{pmatrix}.$$

The shape of this distribution is illustrated by the black contours in Fig. 2a. Here, the observation measures the same aspect as the first component of our state variable:  $\mathcal{H} = [1, 0]$ . As in Fig. 1, assume  $\mathbf{R} = 100$ , so  $\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T + \mathbf{R} \simeq 150.73 + 100.00 = 250.73$ .  $\hat{\mathbf{P}}^b\mathcal{H}^T \simeq [150.73, 109.70]^T$ , and hence  $\hat{\mathbf{K}} = \mathbf{P}^b\mathcal{H}^T(\mathcal{H}\mathbf{P}^b\mathcal{H}^T + \mathbf{R})^{-1} \simeq [0.60, 0.44]^T$ .

For the EnKF, perturbed observations were then generated, denoted by the short vertical lines along the abscissa in Fig. 2a. Eq. (11) was then applied, updating background samples

to their associated perturbed observations, generating analysis samples. For example, the heavy black dot in Fig. 2a was updated to the perturbed observation marked with the “\*”. The resulting analysis sample is the large black dot in Fig. 2b. For the noted sample, the first component of the background state was much less than the mean, and the perturbed observation was greater than the mean background state. The resulting analysis nudged the posterior state toward the mean in both components. Other dots in Fig. 2b denote other updated EnKF member states.

In the EnSRF, the ensemble background mean state  $\sim [47.93, 50.07]^T$  was updated to the mean observed value 58.0 using  $\hat{\mathbf{K}}$  computed above and eq. (16), resulting in a mean analyzed state of  $\sim [53.55, 54.16]$ . As with the EnKF, given the positive observation increment and the positive correlation of the background-error covariances between the two components, both components of the mean state were adjusted upward. EnSRF perturbations from the mean were updated using eq (17) and the reduced gain, here  $\tilde{\mathbf{K}} \simeq 0.613 \hat{\mathbf{K}}$ .

Compare the EnKF and EnSRF random samples of the posterior from Figs. 2b-c and their fitted distribution (black contours) with the correct Bayesian posterior (red contours). The samples from both distributions do not appear to randomly sample the correct posterior. The EnKF and EnSRF posterior distributions are shifted slightly toward lower values in both components. The EnSRF posterior samples preserve the original shape from the prior, though their values are shifted in mean and compressed together. In comparison, the EnKF samples are randomized somewhat through the assimilation of the perturbed observations, and in this case, its distribution is somewhat more diffuse than that of the EnSRF. The EnKF samples appear to overlap more with the correct distribution than the samples from the EnSRF.

Why can’t ensemble-based methods correctly adjust the prior ensemble to the new observations so that the samples reflect a random draw from the Bayesian posterior? The reason is that ensemble-based methods implicitly assume a second-moment closure; that is, the distributions are assumed to be fully described by means and covariances. The example shown above demonstrates that some inaccuracies can be expected in these analyses if indeed there are higher-moment details in these distributions (Lawson and Hansen 2004). Hopefully, highly non-normal distribu-

tions are not frequently encountered, as radically more expensive and unproven techniques than those discussed here may then be required (e.g., Gordon et al. 1993).

*d. Ensemble propagation of the pdf and model-error parameterization*

In real-world applications, background-error covariances cannot simply be estimated at the next assimilation cycle by conducting an ensemble of deterministic forecasts forward from the current cycle's analyses. Because of model deficiencies, even if the true state of the atmosphere is perfectly known, the resulting forecast will be imperfect:  $\mathbf{x}_{(t+1)}^t = \mathcal{M}(\mathbf{x}_{(t)}^t) + \eta$ , where here we denote the time index in parentheses and  $\mathcal{M}$  is again the nonlinear forecast operator. Let's first assume that our forecast model is unbiased  $\langle \eta \rangle = 0$ , again with model-error covariance  $\langle \eta \eta^T \rangle = \mathbf{Q}$ . In practice, The assumption of no bias is probably not justified, and if the bias can be determined, the forecasts ought to be corrected for this bias (Dee and Todling 2000, Evensen 2003) or more ideally, the forecast model ought to be improved. In any case, consider the error covariance at the next assimilation time. Assume again that forecast error due to initial-condition uncertainty and model error are uncorrelated  $\langle (\mathcal{M}(\bar{\mathbf{x}}_{(t)}^a) - \mathcal{M}(\mathbf{x}_{(t)}^t)) \eta^T \rangle = 0$ , and assume linearity of the error growth  $\mathcal{M}(\bar{\mathbf{x}}_{(t)}^a) - \mathcal{M}(\mathbf{x}_{(t)}^t) \simeq \mathbf{M}(\bar{\mathbf{x}}_{(t)}^a - \mathbf{x}_{(t)}^t)$ . Then the true background-error covariance at the next assimilation time is

$$\begin{aligned} \langle (\bar{\mathbf{x}}_{(t+1)}^b - \mathbf{x}_{(t+1)}^t) (\bar{\mathbf{x}}_{(t+1)}^b - \mathbf{x}_{(t+1)}^t)^T \rangle &= \langle (\mathcal{M}(\bar{\mathbf{x}}_{(t)}^a) - \mathcal{M}(\mathbf{x}_{(t)}^t) - \eta) (\mathcal{M}(\bar{\mathbf{x}}_{(t)}^a) - \mathcal{M}(\mathbf{x}_{(t)}^t) - \eta)^T \rangle \\ &\simeq \langle \mathbf{M}(\bar{\mathbf{x}}_{(t)}^a - \mathbf{x}_{(t)}^t) (\bar{\mathbf{x}}_{(t)}^a - \mathbf{x}_{(t)}^t)^T \mathbf{M}^T \rangle + \langle \eta \eta^T \rangle \\ &= \mathbf{M} \mathbf{P}_{(t)}^a \mathbf{M}^T + \mathbf{Q} \end{aligned} \tag{19}$$

where  $\mathbf{M}$  is again the Jacobian of the nonlinear operator. Consider what happens when covariances are estimated directly from an ensemble of forecasts propagated forward from an ensemble of  $i = 1, \dots, m$  analyses using the fully nonlinear forecast model

$$\mathbf{x}_{i(t+1)}^b = \mathcal{M}(\mathbf{x}_{i(t)}^a), \tag{20}$$

Calculating the expected covariance, we get

$$\begin{aligned}
\left\langle \left( \mathbf{x}_{i(t+1)}^b - \bar{\mathbf{x}}_{(t+1)}^b \right) \left( \mathbf{x}_{i(t+1)}^b - \bar{\mathbf{x}}_{(t+1)}^b \right)^T \right\rangle &= \left\langle \left( \mathcal{M}(\mathbf{x}_{i(t)}^a) - \mathcal{M}(\bar{\mathbf{x}}_{(t)}^a) \right) \left( \mathcal{M}(\mathbf{x}_{i(t)}^a) - \mathcal{M}(\bar{\mathbf{x}}_{(t)}^a) \right)^T \right\rangle \\
&\simeq \left\langle \mathbf{M} \left( \mathbf{x}_{i(t)}^a - \bar{\mathbf{x}}_{(t)}^a \right) \left( \mathbf{x}_{i(t)}^a - \bar{\mathbf{x}}_{(t)}^a \right)^T \mathbf{M}^T \right\rangle \\
&\simeq \mathbf{M} \hat{\mathbf{P}}_{(t)}^a \mathbf{M}^T.
\end{aligned} \tag{21}$$

Comparing (19) and (21), it is apparent that an ensemble of analyses that are simply propagated forward with the nonlinear forecast model will have too small an expected amount of spread, missing the extra model-error covariance  $\mathbf{Q}$ . Let us define some hypothetical set of background forecasts at time  $t + 1$  that *do* have the correct covariance, i.e., define  $\check{\mathbf{x}}_{i(t+1)}^b$  such that  $\left\langle \left( \check{\mathbf{x}}_{i(t+1)}^b - \bar{\check{\mathbf{x}}}_{(t+1)}^b \right) \left( \check{\mathbf{x}}_{i(t+1)}^b - \bar{\check{\mathbf{x}}}_{(t+1)}^b \right)^T \right\rangle = \mathbf{M} \hat{\mathbf{P}}_{(t)}^a \mathbf{M}^T + \mathbf{Q}$ . Such an ensemble is possible if we add noise to our existing ensemble:

$$\check{\mathbf{x}}_{i(t+1)}^b = \mathbf{x}_{i(t+1)}^b + \xi_i, \tag{22}$$

where  $\left\langle \xi_i \xi_i^T \right\rangle = \mathbf{Q}$ ,  $\left\langle \xi_i \right\rangle = 0$ , and  $\left\langle \mathbf{x}_{i(t+1)}^b \xi_i^T \right\rangle = 0$ .

Several methods have been proposed for incorporating noise into the ensemble of forecasts so that they account for model error. First, the forecast model could be stochastic-dynamic instead of deterministic, with additional terms in the prognostic equations to represent interactions with unresolved scales and/or mis-parameterized effects; in essence,  $\mathcal{M}$  is changed so that the ensemble of forecasts integrates random noise in addition to the deterministic forecast dynamics, as in (4). Over an assimilation cycle, this additional variance added to the ensemble as a result of integrating noise should be designed to increase the covariance by the missing  $\mathbf{Q}$ . Another possibility is that one may choose to run a forecast model without integrating noise but to add noise to each member at the data assimilation time so as to increase the ensemble variance appropriate to the missing  $\mathbf{Q}$ . Third, it may be possible to use a multi-model ensemble to estimate covariances, or to achieve satisfactory results by inflating the deviations of ensemble members about their mean.

Little work has yet been done on the first of these three approaches. Buizza et al. (1999) demonstrated a simple technique for integrating noise to account for deterministic sub-gridscale

parameterizations. Under their methodology, the parameterized terms in the prognostic equations were multiplied by a random number. Penland (2003) outlines a more general approach for integrating system noise in numerical models. To date, however, a comprehensive noise integration scheme has not yet been demonstrated in an operational weather prediction model. Palmer (2001) discusses the potential appeal of such an approach.

The second general approach is to augment the ensemble-estimated model of covariances during the update step with noise representing the missing model error covariances. Mitchell and Houtekamer (2000) describe one such approach.

A third approach, use of multiple forecast models for generating the ensemble of background forecasts (e.g., Houtekamer et al. 1996b, Harrison et al. 1999, Evans et al. 2000, Ziehmann 2000, Richardson 2000, Hou et al. 2001), is appealing for its simplicity. A wider range of forecasts is typically generated when different weather forecast models are used to forecast the evolution of different ensemble members. Unfortunately, it is not clear whether or not the differences between members are actually representative of model errors; initial experimentation has shown that the multi-model ensembles tend to produce unrealistic estimates of error covariances. Forecast errors at larger scales ought to be mostly in balance, but when estimated from multi-model ensembles, preliminary results suggest that the errors can be greatly out of balance, with detrimental effects on the subsequent assimilation (personal communication, M. Buehner). See also Hansen (2002) for a discussion of discussion of the use of multi-model ensembles in data assimilation in a simple model.

A last approach is to modify the observation- or background-error covariances in some manner so they draw more to the observations. Pham (2001) proposes reducing  $\mathbf{R}$  with a “forgetting factor” to achieve this. Another approach is “covariance inflation,” discussed in Anderson and Anderson (1999). Ensemble members’ deviations about their mean are inflated by an amount  $r$  (slightly greater than 1.0) before the first observation is assimilated:

$$\mathbf{x}_i^b \leftarrow r(\mathbf{x}_i^b - \bar{\mathbf{x}}^b) + \bar{\mathbf{x}}^b. \quad (23)$$

Here, the operation  $\leftarrow$  denotes a replacement of the previous value of  $\mathbf{x}_i^b$ . Application of a moderate inflation factor has been found to improve the accuracy of assimilations (Hamill et al. 2001, Whitaker and Hamill 2002, Whitaker et al. 2004). Note that inflation increases the spread of the ensemble, but it does not change the subspace spanned by the ensemble. Hence, if model error projects into a substantially different subspace, this parameterization may not be effective.

#### *e. Covariance localization*

In ensemble assimilation methods, the accuracy of error statistics is especially important. Unlike 3D-Var, the effects of a mis-specification of error statistics can affect the analysis-error covariance, which is then propagated forward in time. Hence, if the analysis errors are underestimated in one cycle, the forecast errors may be underestimated in the following cycle, underweighting the new observations. The process can feed back on itself, the ensemble assimilation method progressively ignoring observational data more and more in successive cycles, leading eventually to a useless ensemble. This is known as *filter divergence* (e.g., Maybeck 1979, p. 337, Houtekamer and Mitchell 1998). For the ensemble-based methods, filter divergence can be induced by many causes. One of the most crucial is to model background-error covariance realistically (Hamill et al. 2001). As discussed in the previous section, an adequate parameterization of model error is necessary in all but perfect model simulations. However, filter divergence can occur even in simulations where the forecast model is perfect, for background-error covariances are typically estimated imperfectly from small ensembles. While more ensemble members would be desirable to reduce the sampling error in estimating background-error covariances, more members means more computational expense.

One common algorithmic modification to improve error covariance estimates from ensembles is *covariance localization*. The covariance estimate from the ensemble is multiplied point by point with a correlation function that is 1.0 at the observation location and zero beyond some prespecified distance. Houtekamer and Mitchell (1998), Ott et al. (2003) and Evensen (2003) use a cutoff radius so that observations are not assimilated beyond a certain distance from the grid point.



A preferable approach is to use a correlation function that decreases monotonically with increasing distance (Houtekamer and Mitchell 2001). Mathematically, to apply covariance localization, the Kalman gain  $\hat{\mathbf{K}} = \hat{\mathbf{P}}^b \mathcal{H}^T (\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T + \mathbf{R})^{-1}$  is replaced by a modified gain

$$\hat{\mathbf{K}} = (\rho_S \circ \hat{\mathbf{P}}^b) \mathcal{H}^T (\mathcal{H} (\rho_S \circ \hat{\mathbf{P}}^b) \mathcal{H}^T + \mathbf{R})^{-1}, \quad (24)$$

where the operation  $\rho_S \circ$  in (24) denotes a Schur product (an element-by-element multiplication) of a correlation matrix  $\mathbf{S}$  with local support with the covariance model generated by the ensemble. For horizontal localization, one such correlation matrix can be constructed using an approximately Gaussian-shaped function described in Gaspari and Cohn (1999). The Schur product of matrices  $\mathbf{A}$  and  $\mathbf{B}$  is a matrix  $\mathbf{C}$  of the same dimension, where  $c_{ij} = a_{ij} b_{ij}$ . When covariance localization is applied to smaller ensembles, it can actually result in more accurate analyses than would be obtained from larger ensembles without localization (Houtekamer and Mitchell 2001). Mathematically, localization increases the effective rank of the background-error covariances (Hamill et al. 2001). In the extreme, if the correlation matrix  $\mathbf{S}$  were the identity matrix, the covariance model would consist of grid points with variances and zero covariance and the rank of the covariance matrix after localization would increase from  $m - 1$  to  $n$ , the dimension of the state vector. In practice, such an extreme localization would harm the quality of the analysis, destroying the mass-wind balance (Mitchell and Houtekamer 2002, Lorenc 2003) and prohibiting the observation from changing the analysis at nearby grid points. Hence, broader localizations are typically used. Generally, the larger the ensemble, the broader the optimum correlation length scale of the localization function (Houtekamer and Mitchell 2001, Hamill et al. 2001). See Whitaker et al. (2004) and Houtekamer et al. (2004) for examples of ensemble assimilations that also include a vertical covariance localization.

As a concrete example of horizontal covariance localization, consider Fig. 4. This used the same data set used in Fig. 2, a global ensemble-data assimilation scheme utilizing only sea-level pressure observations (Whitaker et al. 2004). Figure 4a provides a map of sea-level pressure correlations at grid points around the Northern Hemisphere with a grid point in the western Pacific Ocean on 0000 UTC 14 December 2001. When directly estimated using the 25-member

ensemble subsampled from the 200-member ensemble (Fig. 4b), correlations for grid points in the region around the observation are positive. The shape of the correlation function was anisotropic, with positive correlations generally limited to a region east of the axis of the cyclone. Background errors for regions in the eastern Pacific and near the Greenwich meridian also appeared to be highly correlated with background errors at the observation location. However, when the correlations are estimated from a 200-member ensemble, it's apparent that these distant correlations in the 25-member ensemble were caused by the limited sample size. The errors in the eastern Pacific and along the Greenwich meridian were not dynamically interconnected with the errors in the western Pacific. When the covariance localization function (Fig. 4c) was applied to the 25-member ensemble, the resulting correlation model (Fig. 4d) more closely resembles that from the larger ensemble.

In applying the covariance localization, distant grid points are forced to be statistically independent. Should they be? As a thought experiment, consider a two-member ensemble. Dynamically, there is no a priori reason to expect that, say, the growth of spread over Japan is dynamically interconnected to the growth of spread over Africa, and neither interconnected with the growth of differences over South America. This two-member ensemble may identify many distinct regions where rapid growth of differences is occurring, regions but with a covariance model estimated from only 2 members, it assumes they are all intimately coupled. Covariance localization is thus a heuristic attempt to modify the model of background-error covariances so that a limited-size ensemble will not represent distant, distinct features as dynamically interrelated when in fact they only appear to be due to limited sample size. If indeed distant regions are in fact dynamically coupled, the localization will cause the loss of this information. The effect on the data assimilation will be that observations will not be able to change the analysis and reduce the analysis-error variance in distant regions; local observations will have to be relied upon instead. This is judged to be less detrimental than the opposite, to let observations affect distant regions when this is inappropriate.

## 5. DISCUSSION

### *a. Major research questions*

Researchers are just beginning to test ensemble-based atmospheric data assimilation methods in full numerical weather prediction modeling systems using real observations. From these and other studies, we can make an educated guess at what some of the major issues that will need to be resolved before operational implementation is practical.

As discussed previously, in order to ensure a high-quality analysis, great care must be taken to ensure that the error-covariance models are realistic in ensemble methods. These methods cycle the covariance estimates. Thus, for example, if observations errors are assumed to be unbiased and independent but in fact are biased or correlated (Liu and Rabier 2003), these errors will cause the analysis-error covariance to be misestimated, later affecting the subsequent background-error estimates and subsequent fit to the observations. Accurate estimation of model error in particular is likely to be crucial, as was discussed in the previous section 4d.

Practically, an ensemble-based assimilation method ought to be self-correcting, able to detect when the system is not appropriately fitting the observations. Theoretically, this can be done by monitoring the innovation statistics ( $\mathbf{y} - \mathcal{H}(\mathbf{x}^b)$ ), which ought to be white noise with zero mean and covariance ( $\mathcal{H}\hat{\mathbf{P}}^b\mathcal{H}^T + \mathbf{R}$ ) (Maybeck sect. 5.4, Dee 1995). The influence of model error can then be increased or decreased so the innovation statistics have the correct properties (Mitchell and Houtekamer 2000).

Other problems may be more subtle. For instance, initial tests (e.g., Houtekamer et al. 2004) suggest that when many observations are frequently assimilated, the errors due to chaotic effects may not grow rapidly after the analysis, as expected. The reasons for this are not yet fully apparent. It is known that the more observations that are assimilated, the whiter and more random are the analysis errors (Hamill et al. 2003); consequently, it may take longer than the time between updates for the dynamics to organize the perturbations into growing structures. The randomization of analysis errors may also be exacerbated by the addition of random model error to the background forecasts or because the computational costs require the use of reduced-resolution models with unrealistically slow error growth characteristics.

A final major concern is the computational expense. The cost of most ensemble methods scales as the number of observations times the dimension of the model state times the number of ensemble members. In the coming years, observations will increase in number faster than computer power will grow. It may be possible to mitigate this problem in one of several ways. Perhaps computations can be speeded up through parallelization (Houtekamer and Mitchell 2001, Keppenne and Rienecker 2002), the method can be cast in a variational framework where the costs do not scale with the number of observations (Hamill and Snyder 2000), or many high-density observations can be combined into fewer “super-observations” (Lorenc 1981).

*b. Comparisons with 4D-Var*

An important question is whether, for a given amount of computer time, a better analysis could be produced by an ensemble-based assimilation or by the current state-of-the art, four-dimensional variational analysis (4D-Var; Le Dimet and Talagrand 1986, Courtier et al. 1994, Rabier et al. 1998, 2000). Such direct comparisons of ensemble assimilation methods and 4D-Var in realistic scenarios have yet to be performed and ideally should wait until ensemble methods are more mature.

Some intelligent guesses can be made regarding their relative advantages and disadvantages; for another view, see Lorenc (2004). Ensemble-based methods are much easier to code and maintain, for neither a tangent linear nor an adjoint model of the forecast model is required, as they are with 4D-Var. Ensemble-based methods produce an ensemble of possible analysis states, providing information on both the mean analysis and its uncertainty. Consequently, the ensemble of analysis states can be used directly to initialize ensemble forecasts without any additional computations.

Another advantage is that if the analysis uncertainty is very spatially inhomogeneous and time dependent, in ensemble-based methods this information will be fed through the ensemble from one assimilation cycle to the next. In comparison, in 4D-Var, the assimilation typically starts at each update cycle with the same stationary model of error statistics. Hence, the influence of observations may be more properly weighted in ensemble-based methods than in 4D-

Var. Ensemble-based methods also provide a direct way to incorporate the effects of model imperfections directly into the data assimilation. In comparison, in current operational implementations of 4D-Var, the forecast model dynamics are a strong constraint (Courtier et al. 1994; but see Bennett et al. 1996 for a possible alternative). If the forecast model used in 4D-Var does not adequately represent the true dynamics of the atmosphere, model error may be large, and 4D-Var may fit a model trajectory that was significantly different than the trajectory of the real atmosphere during that time window.

Ensemble-based techniques may have disadvantages relative to 4D-Var, some that will only be discovered through experimentation. Ensemble-based techniques are likely to be at least as computationally expensive as 4D-Var, and perhaps significantly more expensive when there are an overwhelmingly large number of observations. Ensemble approaches may be difficult to apply in limited-area models because of the lateral boundary conditions, and the method is very sensitive to misestimation of the error covariances.

### *c. Applications of ensemble-based assimilation methods*

Ensemble data assimilation techniques offer the potential of generating calibrated analyses that may be useful for a variety of applications. Anderson (2001) showed that the ensemble techniques can be used for parameter estimation. Hamill and Snyder (2002) showed that ensemble assimilation techniques facilitate the calculation of regions where adaptive observations are necessary. Snyder and Zhang (2003), Zhang et al. (2003), and Dowell et al. (2004) demonstrate the feasibility of ensemble filters for mesoscale data assimilation of radar observations. Reichle et al. (2002ab, 2003) apply ensemble filters to estimation of soil moisture. Hamill et al. (2003) demonstrate how analysis-error covariance singular vectors, the most rapidly growing forecast structures consistent with analysis errors, can be diagnosed using ensemble filters.

## **6. CONCLUSIONS**

This manuscript presented a brief tutorial of ensemble-based data assimilation. The technique is being explored by a rapidly growing number of researchers as a possible alternative to

other atmospheric data assimilation techniques such as three- and four-dimensional atmospheric data assimilation. The technique is appealing for its comparative algorithmic simplicity and its ability to explicitly deal with model error. Testing of ensemble filters has progressed rapidly over the past few years from perfect-model experiments in toy systems to the assimilation of real observations into global models. Recent results are both suggestive of the potential and hint that substantial continued development is necessary for these methods to be competitive with or superior to the existing state-of-the-art techniques.

## 7. ACKNOWLEDGMENTS

Deszo Devenyi (NOAA/FSL), Chris Snyder (NCAR/MMM), Jeff Whitaker (NOAA/CDC), Ryan Torn (U. Washington), Brian Etherton (UNC/Charlotte), Sharanya Majumdar (U. Miami), Jim Hansen (MIT) and three anonymous reviewers are thanked for their informal reviews of early versions of this manuscript. This paper was originally prepared for ECMWF's 2002 Predictability Workshop. The many participants who offered interesting questions and constructive criticism at this workshop are thanked as well. Preparation of this manuscript was supported by National Science Foundation grants ITR-0112715 and ATM-0205612.

## 8. APPENDIX: SERIAL PROCESSING OF NON-INDEPENDENT OBSERVATIONS IN THE ENSEMBLE-BASED ASSIMILATION

As previously outlined, if observations have independent errors, then they can be processed serially, which may be of computational advantage. If the observations are not independent, the EnKF update equation (19) can be transformed so that serial processing of observations can occur.

Recall that  $\mathbf{y}_t = \mathcal{H}(\mathbf{x}_t^\dagger) + \epsilon$ , where  $\epsilon \sim N(0, \mathbf{R})$ , where  $\mathbf{R}$  is no longer assumed to be diagonal. However,  $\mathbf{R}$  is symmetric and positive definite, so it has a decomposition of the form  $\mathbf{R} = \mathbf{Q}_R \Lambda_R \mathbf{Q}_R^T$ , where  $\mathbf{Q}_R$  is a unitary matrix with properties that  $\mathbf{Q}_R \mathbf{Q}_R^T = \mathbf{I}$  and  $\mathbf{Q}_R^T = \mathbf{Q}_R^{-1}$  (here  $\mathbf{Q}_R$  does not denote model error).  $\Lambda_R$  is a diagonal matrix of associated eigenvalues.

Let's denote a pseudo-observation  $\tilde{\mathbf{y}} = \mathbf{Q}_R^T \mathbf{y}$ , or alternately,  $\mathbf{y} = \mathbf{Q}_R \tilde{\mathbf{y}}$ . Then  $\tilde{\mathbf{y}} = \mathbf{Q}_R^T \mathcal{H}(\mathbf{x}_t^\dagger) + \mathbf{Q}_R^T \epsilon$ . Hence

$$\langle \mathbf{Q}_R^T \epsilon (\mathbf{Q}_R^T \epsilon)^T \rangle = \mathbf{Q}_R^T \langle \epsilon \epsilon^T \rangle \mathbf{Q}_R = \Lambda_R. \quad (\text{A1})$$

Define  $\tilde{\mathcal{H}} = \mathbf{Q}_R^T \mathcal{H}$ , or equivalently  $\mathcal{H} = \mathbf{Q}_R \tilde{\mathcal{H}}$ . Substituting this definition of  $\mathcal{H}$  and  $\mathbf{y}$  into the EnKF update equation (25), we get

$$\begin{aligned} \mathbf{x}_i^a &= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \mathcal{H}^T (\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T + \mathbf{R})^{-1} (\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i^b)) \\ &= \mathbf{x}_i^b + \hat{\mathbf{P}}^b (\mathbf{Q}_R \tilde{\mathcal{H}})^T (\mathbf{Q}_R \tilde{\mathcal{H}} \hat{\mathbf{P}}^b \tilde{\mathcal{H}}^T \mathbf{Q}_R^T + \mathbf{Q}_R \Lambda_R \mathbf{Q}_R^T)^{-1} (\mathbf{Q}_R \tilde{\mathbf{y}}_i - \mathbf{Q}_R \tilde{\mathcal{H}}(\mathbf{x}_i^b)) \\ &= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \tilde{\mathcal{H}}^T \mathbf{Q}_R^T \mathbf{Q}_R (\tilde{\mathcal{H}} \hat{\mathbf{P}}^b \tilde{\mathcal{H}}^T + \Lambda_R)^{-1} \mathbf{Q}_R^T \mathbf{Q}_R (\tilde{\mathbf{y}}_i - \tilde{\mathcal{H}}(\mathbf{x}_i^b)) \\ &= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \tilde{\mathcal{H}}^T (\tilde{\mathcal{H}} \hat{\mathbf{P}}^b \tilde{\mathcal{H}}^T + \Lambda_R)^{-1} (\tilde{\mathbf{y}} - \tilde{\mathcal{H}}(\mathbf{x}_i^b)). \end{aligned} \quad (\text{A2})$$

Thus, given a batch of observations with correlated errors and known observation-error covariance matrix  $\mathbf{R}$  for these observations, one determines the eigenvectors  $\mathbf{Q}_R$  and eigenvalues  $\Lambda_R$  of  $\mathbf{R}$ , forms the transformed perturbed observations  $\tilde{\mathbf{y}}$  and operator  $\tilde{\mathcal{H}}$  and then solves the last line of (A21) can be used to serially process observations. See Kaminski et al. (1971) for an essentially equivalent algorithm using a Cholesky decomposition.

## REFERENCES

- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.
- , 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.
- , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634-642.
- Andrews, A., 1968: A square-root formulation of the Kalman covariance equations. *AIAA J.*, **6**, 1165-1168.
- Bennett, A. F., B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Met. Atmos. Phys.*, **60**, 165-178.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part 1: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.
- Blanchet, I., C. Frankignoul, and M. A. Cane, 1997: A comparison of adaptive Kalman filters for a tropical ocean model. *Mon. Wea. Rev.*, **125**, 40-58.
- Bouttier, F., 1994: A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.*, **122**, 2376-2390.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.
- Casella, G., and R. L. Berger, 1990: *Statistical inference*. Duxbury Press, 650 pp.
- Cohn, S. E., and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757-1785.
- , 1997: An introduction to estimation theory. *J. Meteor. Soc. Jap.*, **75(1B)**, 257-288.



- Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press. 457 pp.
- , 1992: Estimating model-error covariances for applications to atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 1735-1746.
- , 1993: Estimating observation error statistics for atmospheric data assimilation. *Annales Geophysicae*, **11**, 634-647.
- , 1997: Atmospheric data assimilation. *J. Meteor. Soc. Jap.*, **75(1B)**, 319-329.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128-1145.
- Dee, D. P., and R. Todling, 2000: Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Wea. Rev.*, **128**, 3268-3282.
- Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and thermodynamic retrievals in the 17 May 1981 Arcadia, Oklahoma supercell: ensemble Kalman filter experiments. *Mon. Wea. Rev.*, in press.
- Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: theory. *Mon. Wea. Rev.*, **122**, 703-713.
- Ehrendorfer, M., 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: applications. *Mon. Wea. Rev.*, **122**, 714-728.
- Evans, R. E., M. S. J. Harrison, and R. J. Graham, 2000: Joint medium-range ensembles from the Met. Office and ECMWF systems. *Mon. Wea. Rev.*, **128**, 3104-3127.
- Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17905-17924.
- , 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143-10162.

- , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.
- , 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343-367.
- Farrell, B. F., and P. J. Ioannou, 2001: State estimation using a reduced- order Kalman filter. *J. Atmos. Sci.*, **58**, 3666-3680.
- Fisher, M., 1998: Development of a simplified Kalman filter *ECMWF Research Department Technical Memorandum 260*. European Centre for Medium-Range Weather Forecasts. 16 pp. Available from Library, ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.
- Gardiner, C. W., 1985: *Handbook of Stochastic Methods (2nd ed.)*. Springer, 444 pp.
- Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.
- Gauthier, P., P. Courtier, and P. Moll, 1993: Assimilation of simulated lidar data with a Kalman filter. *Mon. Wea. Rev.*, **121**, 1803-1820.
- Gelb, A. (ed.), 1974: *Applied optimal estimation*. MIT Press, 374 pp.
- Ghil, M., 1989: Meteorological data assimilation for oceanography. Part 1: description and theoretical framework. *Dyn. Atmos. Oceans*, **13**, 171-218.
- , and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141-266.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings - F*, **140**, 107-113.
- Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*, John Wiley and Sons Press, 477 pp.
- Hamill, T. M., and C. Snyder, 2000. A hybrid ensemble Kalman filter / 3d-variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

- , J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776-2790.
- , and C. Snyder, 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552-1572.
- , ———, and R. E. Morss, 2002: Analysis-error statistics of a quasigeostrophic model using three-dimensional variational assimilation. *Mon. Wea. Rev.*, **130**, 2777-2790.
- , ———, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741-1758.
- Hansen, J. A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.*, **130**, 2373-2391.
- Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Quart. J. Roy. Meteor. Soc.*, **125**, 2487-2515.
- Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning*. Springer, 533 pp.
- Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance-reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718-1728.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX-98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73-91.
- Houtekamer, P. L., 1993: Global and local skill forecasts. *Mon. Wea. Rev.*, **121**, 1834-1846.
- , L. Lefaivre, and J. Derome, 1996a: The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability, Vol II*, Reading, United Kingdom, 121-146. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom].
- , J. Derome, H. Ritchie, and H. L. Mitchell, 1996b: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.
- , and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

- , and ———, 1999: Reply to comment on “Data assimilation using an ensemble Kalman filter technique.” *Mon. Wea. Rev.*, **127**, 1378-1379.
- , and ———, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.
- , ———, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2004: Atmospheric data assimilation with the ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.*, in review.
- Jazwinski, A. H., 1970: Stochastic processes and filtering theory. Academic Press, 376 pp.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the AMSE- Journal of Basic Engineering*. **82D**, 35-45.
- , and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *Transactions of the AMSE-Journal of Basic Engineering*. **83D**, 95-108.
- Kaminski, P. G., A. E. Bryson, Jr., and S. F. Schmidt, 1971: Discrete square root filtering: a survey of current techniques. *IEEE Transactions on Automatic Control*, **AC-16**, 727-736.
- Keppenne, C. L., 2000: Data assimilation into a primitive equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971-1981.
- , and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, **130**, 2951-2965.
- Kistler, R., and Coauthors, 2001: The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-268.
- Lacarra, J. F., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus*, **40A**, 81-95.
- Lawson, G., and J. A. Hansen, 2003: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, in press.

- Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, , **38A**, 97-110.
- Leith, C. E., 1983: Predictability in theory and practice. Chapter 13 in *Large-scale Dynamical Processes in the Atmosphere*. B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 397 pp.
- Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part 1: theory and schemes. *Mon. Wea. Rev.*, **127**, 1385-1407.
- , 2002: On the mapping of multivariate geophysical fields: sensitivities to size, scales, and dynamics. *J. Atmos. Oceanic Technol.*, **19**, 1602-1637.
- Li, Z. and I. M. Navon, 2001: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Quart. J. Roy. Meteor. Soc.*, **127**, 661-683.
- Liu, Z.-Q., and F. Rabier, 2003: The potential of high-density observations for numerical weather prediction: a study with simulated observations. *Quart. J. Roy. Meteor. Soc.*, **129**, 3013-3035.
- Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.
- , 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194.
- , 2003: The potential of the ensemble Kalman filter for NWP. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183-3204.
- Maybeck, P. S., 1979: *Stochastic models, estimation, and control. Volume 1*. Academic Press, 423 pp.
- Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167-194.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.
- , and ———, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791-2808.

- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.
- Ott, E., B. R. Hunt, I. Szunyogh, M. Corazza, E. Kalnay, D.J. Patil, J. A. Yorke, A. V. Zimin, and E. J. Kostelich, 2003: Exploiting local low dimensionality of the atmospheric dynamics for efficient ensemble Kalman filtering. *Mon. Wea. Rev.*, submitted. Available from eo4@umail.umd.edu .
- Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279-304.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747-1763.
- Penland, C., 2003: A stochastic approach to nonlinear dynamics: a review. *Bull. Amer. Meteor. Soc.*, (Electronic supplement to "Noise out of chaos and why it won't go away, *Bull. Amer. Meteor. Soc.*, , **84**, 921-925).
- Pham, D. T., 2001 : Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194-1207.
- Potter, J., 1964: W matrix augmentation. *M. I. T. Instrumentation Laboratory Memo SGA 5-64*, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1-39.
- , H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.
- Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103-114.

- , J. P. Walker, R. D. Koster, and P. R. Houser, 2002b: Extended vs. ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.*, **3**, 728-740.
- , and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.*, accepted. Available from reichle@janus.gsfc.nasa.gov .
- Richardson, D. S., 2000: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.*, **127**, 1847-1864.
- Sardeshmukh, P. D., C. Penland, and M. Newman, 2001: Rossby waves in a stochastically fluctuating medium. In *Stochastic Climate Models*, ed. P. Imkeller and J.-S. von Storch, Progress in Probability, Birkhauser, Basel, pp. 369-384.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663-1677.
- Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Jap.*, **75(1B)**, 191-209.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485-1490.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.
- , and ———, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.
- van Leeuwen, P.J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique." *Mon. Wea. Rev.*, **127**, 1374-1377.
- Verlaan, M. and A. W. Heemink, 2001: Nonlinearity in data assimilation applications. A practical method for analysis. *Mon. Wea. Rev.*, **129**, 1578-1589.
- Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140-1158.

- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.
- , G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, in press. Available from [jeffrey.s.whitaker@noaa.gov](mailto:jeffrey.s.whitaker@noaa.gov).
- Wu, W.-S., M. Iridell, S. Saha, and P. Caplan, 1997: Changes to the 1997 NCEP operational MRF model analysis/forecast system. *Technical Procedures Bulletin 443*, NCEP, National Weather Service, Office of Meteorology, Programs and Plans Division, Silver Spring MD, 20910. Available from <http://www.nws.noaa.gov/om/tpb/indexb.htm>.
- , R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905-2916.
- Zhang, F., C. Snyder, and J. Sun, 2003: Impact of initial estimate and observations on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, in review. Available at <http://www.met.tamu.edu/personnel/faculty/fzhang/PAPERS/ZhangSnyderSun2003MWR.pdf>.
- Ziehmann, C, 2000: Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52a**, 280-299.



## FIGURE CAPTIONS

Figure 1. Example of Bayesian data assimilation update. Here the model state is two dimensional and a single observation is assimilated. This observation measures the same variable as the first component of the model state. (a) Probability density for prior joint and marginal distributions (solid) and sample observation distribution (dashed). The three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively. (b) Probability density for posterior distributions. Contours levels set as in (a).

Figure 2. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa.

Figure 3. Illustration of the EnKF and EnSRF with a two-dimensional state variable and observations observing the same as  $x_{(1)}^b$ . (a) Random samples (black dots) from the probability distribution in (1), and the original prior pdf, contoured in red. Implied bivariate normal probability background distribution estimated from the sample ensemble contoured in black, and the observation sampling distribution (dashed). Solid vertical lines along abscissa denote individual perturbed observations sampled from this distribution. The one large black dot and the perturbed observation marked with a star denote the sample discussed in the text. (b) Analyzed samples from the EnKF assimilation scheme (dots), the implied analysis-error bivariate normal distribution from this sample (solid black contours), and the true posterior pdf from Fig. 1 (red). (c) Analyzed samples from EnSRF (dots), implied bivariate normal pdf (solid black contours) and the true posterior pdf (red). In each panel, the three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively.

Figure 4. Illustration of covariance localization. (a) Correlations of sea-level pressure directly estimated from 25-member ensemble with pressure at a point in the western Pacific (colors). Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa. (b)

As in (a), but using 200-member ensemble. (c) Localization function, (d) Correlation estimate from 25-member ensemble after application of covariance localization.

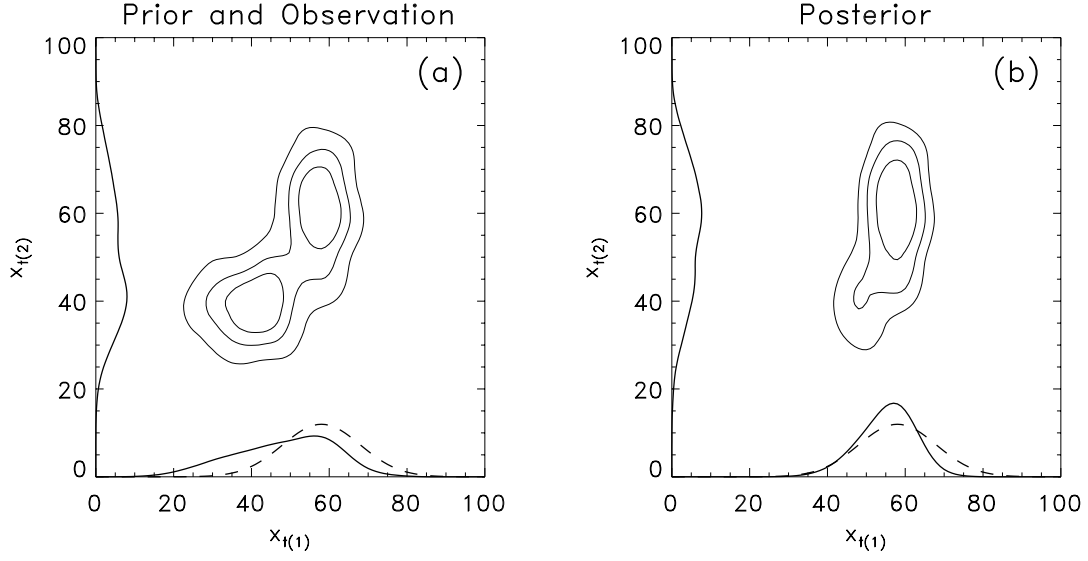


Figure 1. Example of Bayesian data assimilation update. Here the model state is two dimensional and a single observation is assimilated. This observation measures the same variable as the first component of the model state. (a) Probability density for prior joint and marginal distributions (solid) and sample observation distribution (dashed). The three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively. (b) Probability density for posterior distributions. Contours levels set as in (a).

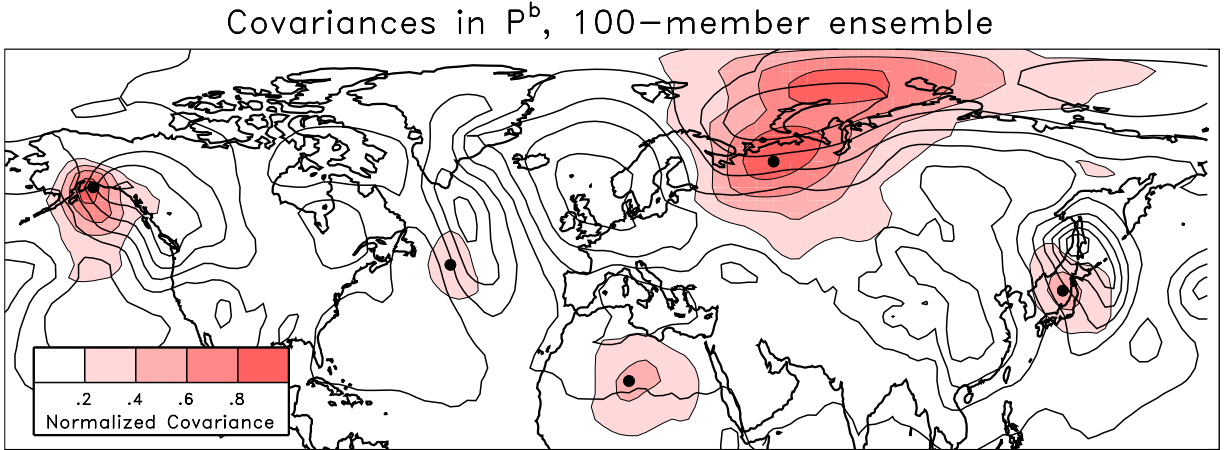


Figure 2. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa.

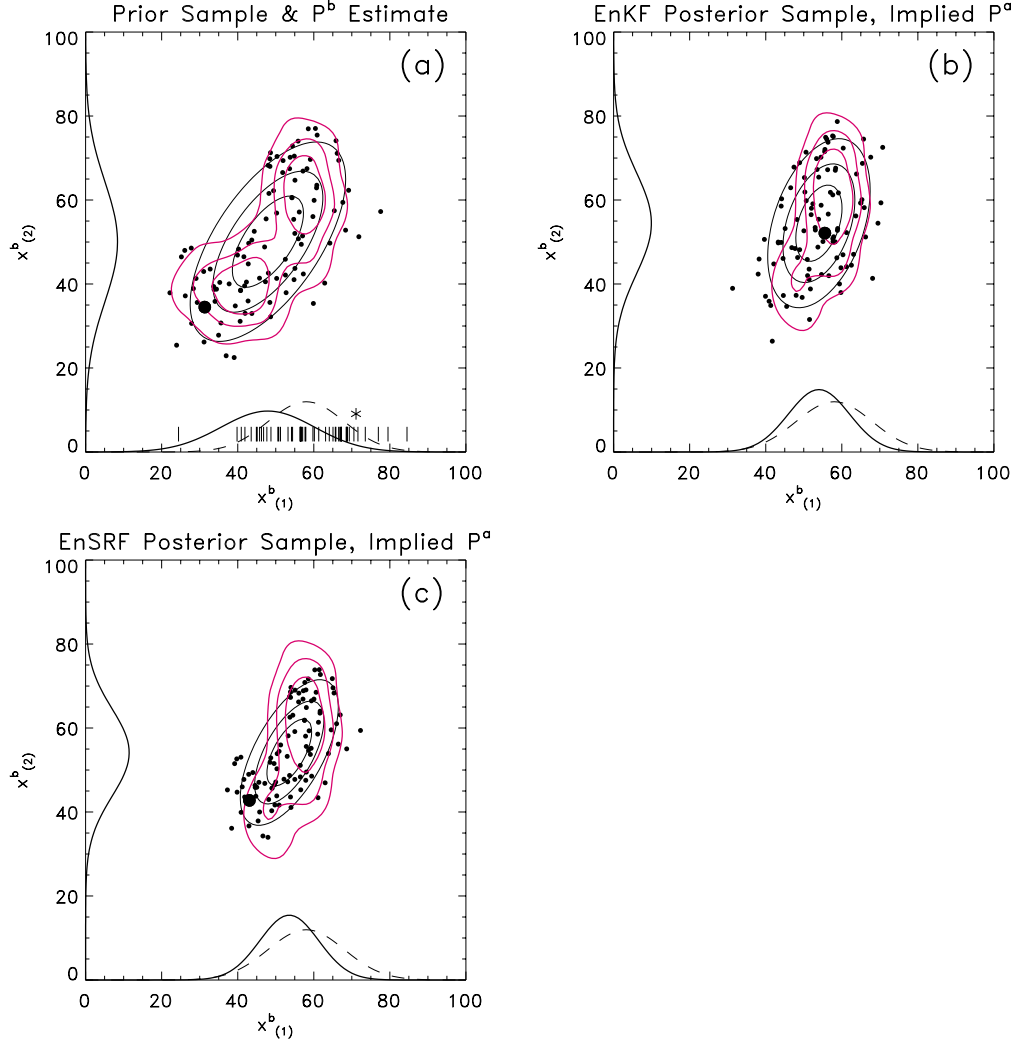


Figure 3. Illustration of the EnKF and EnSRF with a two-dimensional state variable and observations observing the same as  $x^b_{(1)}$ . (a) Random samples (black dots) from the probability distribution in (1), and the original prior pdf, contoured in red. Implied bivariate normal probability background distribution estimated from the sample ensemble contoured in black, and the observation sampling distribution (dashed). Solid vertical lines along abscissa denote individual perturbed observations sampled from this distribution. The one large black dot and the perturbed observation marked with a star denote the sample discussed in the text. (b) Analyzed samples from the EnKF assimilation scheme (dots), the implied analysis-error bivariate normal distribution from this sample (solid black contours), and the true posterior pdf from Fig. 1 (red). (c) Analyzed samples from EnSRF (dots), implied bivariate normal pdf (solid black contours) and the true posterior pdf (red). In each panel, the three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively.

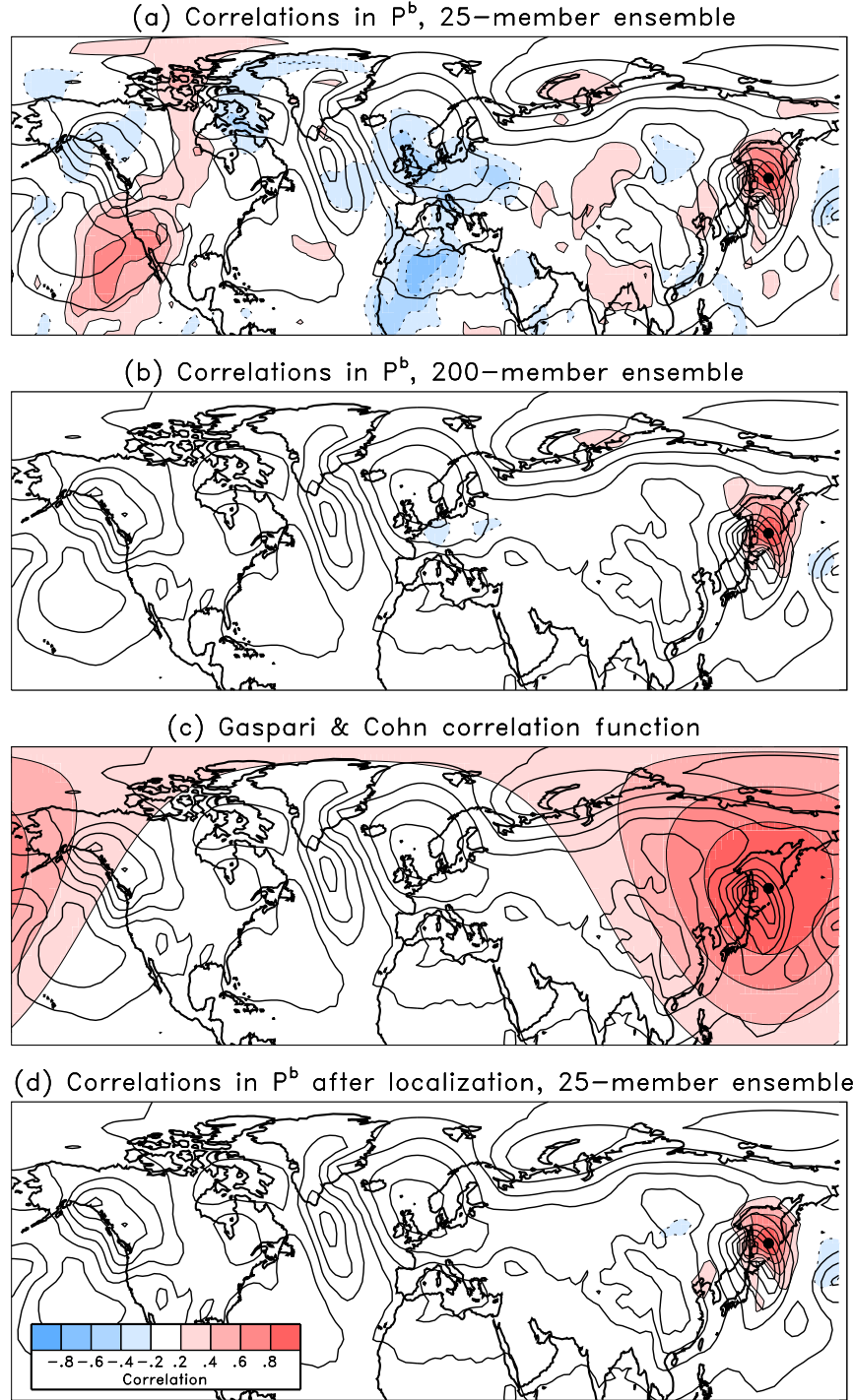


Figure 4. Illustration of covariance localization. (a) Correlations of sea-level pressure directly estimated from 25-member ensemble with pressure at a point in the western Pacific (colors). Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa. (b) As in (a), but using 200-member ensemble. (c) Localization function, (d) Correlation estimate from 25-member ensemble after application of covariance localization.